Advanced Science and Technology Letters
Vol.86 (Ubiquitous Science and Engineering 2015), pp.43-46
http://dx.doi.org/10.14257/astl.2015.86.09

A Study on Search Improvement Using the CST-tree Index

Jung Myoung-jin[1*], Jo Sung-jae[2]

[1] Dept of Education at Dongbang Graduate University, Ph.D in Education of Education
[2] Dept of Education at Dongbang Graduate University Eastern Literature Department

Abstract. The number of web users is increasing due to the dissemination of the internet and the fast communications network. Such popularization of the internet has led to increases in bulk storage data in the web log data. There is a need for a method that analyzes bulk storage data efficiently. This study allowed for improved searches using the CST-tree index by analyzing the web log data that is generated by an user accesses the web server.

1 Introduction

Currently, most of the internet environment is developing into web-based systems. The number of web service users is continuously increasing. Consequently, there is a need for efficient analysis on the bulk storage web log information by either discovering the attack on the web server when the user accesses a large portal site's web server or when attempting to connect with the web mining technology. The existing web long processing methods are not fit to process bulk storage web log at a high speed because it conducted sequential searches. In this study shows innovative analysis and improved search from bulk storage log through the composition of CST-tree indexing vector for the bulk storage web log.

2 Related Research

2.1 B-tree Web Log Mining [1]

The B-tree web+ log mining technique reduced the size of long information by loading the log information, checking overlapping strings and parsing technique per field, using indexes per field unit. Also, the study sought to constitute index information in the field unit based on the B-tree to efficiently reflect the function of the preconditioning process. The technique proposed by this paragraph handles the indexing process for the web log information in order to improve detection of attacks within the bulk-storage web log. Also, indexing-based mining technique and the

ISSN: 2287-1233 ASTL
Copyright © 2015 SERSC

solution were found by reflecting the attack reaction technique every hour[1]. However, since the proposed web log mining is based on B-tree, it is reduced in performance in comparison to the suggested technique based on CST-tree

2.2 Graph Clustering Technique[2]

The graph clustering technique improves the outcome of corporate management and maximizes customers' profit by CRM. The web log type was shown as a graph and the bulk storage data was analyzed with customer characteristic and behavior model. It is data based on the characteristics and the properties with relation between objects. It is superior to the point and transaction data and the data is not limited by an absolute value[2]. However, the proposed graph clustering method performs inferiorly to the proposed technique of the CST-tree base.

2.3 Online Marketing System (ROMS) [3]

The online marketing system (ROMS) technique provides necessary marketing information by continuously collecting and analyzing the processed information of website visitors' behaviors. This technique realized an online marketing system (ROMS) in order to increase the conversion rate of the website. Such online marketing system was created to continuously collect and analyze the user's behavior and use the information to conduct useful online marketing. However, the proposed ROMS technique is inferior in performance than the proposed technique of CST-tree base[3].

2.4 Advertising System Using Web Log [4]

The advertising system that uses web log is a system that services, studies, and performs personalized advertising service through server log analysis to achieve efficient advertising service. This system is constituted of operations of the client and server and it could collect, organize, and analyze the data to fit the propensity of the client. It is a system designed to observe the advertisement transmitting function based on the collected data for each user through the web browser[4]. However, the advertisement system using the web log is inferior in performance than the CST-tree based technique.

2.5 Web Log Analysis Model [5]

The web log analysis model allows the web administrator to extract information through site analysis necessary for web log analysis, use the analyzed data to improve the web environment, and provides high-quality service for website users. In order to achieve such a goal, the web log analysis model studied the analysis principles and techniques of the concept, pattern, and web log of the web log. Using this model, a useful web log analysis model is proposed through the statistical analysis for the

 Copyright © 2015 SERSC

number of visitors, analysis of popular websites, and IP tracing of the visitors[5]. however, the proposed web model is inferior in performance than the CST-tree base technique.

2.6 CST-tree [6]

CST-tree distributes child nodes to successive memory spaces to search a single point to direct to the child node. It is structured to search the number of cache miss that is generated from the size of the T-tree of a premade node. The search technique of binary tree search and CST-tree search are similar. They search by comparing the search key value and node key value intended to be searched.

3 CST-tree Web Log Analysis

In this chapter, the size of the log information was effectively reduced by using the field-unit index technique that applies the parsing technique for each field and the repeating strings by reading log information that is based on CST-tree. Also, by generating index information in field units in the CST-tree structure, the performance of the existing preconditioning function was aimed to be improved.

3.1 Web Log CST-tree

In this study, the CST-tree structure was used to activate an effective indexing structure for all information generated within the log information. The keyword is set as parsing information for each field and the data is set as the index of field information. When the bulk storage web log information is converted into the values of CST-tree base, the overall storage reduces. Also, conducting searches for the strings with index values within the web log information will lead to improved results. Ultimately, the web log information must be based on indexes in order to minimize the time and file size of the preconditioning process.

3.2 Web Log CST-tree Search

It is the structure that allows faster searching of field information in a direct method by using the index numbers from web log index vector values. The index value is searched within the searched information from the CST-tree when searching with basic information. The index number obtained search after search is processed. There is no need to search the same information again after obtaining the index number. The values could be collected with index numbers alone.

Copyright © 2015 SERSC

4 Performance Evaluation Conclusion

In order to compare the performance evaluation of the existing method and the proposed method, the search performance of the **page count** that involved 70,000 lines of web log information, equivalent to 10,000 lines. Conducting the preconditioning process in the method stated above will generate the information of the index key value list and the CST-tree structure. The result of the performance evaluation showed that the CST-tree system is superior, overall, to the B-tree system.

5 Discussion and Conclusion

This study reduced the size of the log information using the CST-tree structure in field units that consider the parsing technique for each field and repeating strings for the web log information. Using the CST-tree structure, the index information for each field within the bulk storage web information was structured and the function of high-speed search and sharing session was given. The result of performance comparison confirmed that the proposed technique is a faster and more effective web log management structure in comparison to the existing technique. By using the CST-tree instead of the B-tree, the use pattern of web users could be understood and an effective web log management structure can be expected by reflecting the function of the web mining system through the pattern of the bulk storage web log.

References

1. Hyung-Woo Lee, Tae-Su Kim, .: High-Speed Search Mechanism based on B-Tree Index Vector for Huge Web Log Mining and Web Attack Detection: Journal of Korea Multimedia Society, Vol.11, No. 11, pp. 1601-1614. November, (2008).
2. Jung-Eun Kim, Jae-Gil Lee, .: Analysis of User Characteristics and Behaviors through Large-Scale Weblog Data: Entrue Journal of Information Technology, Vol.12, No. 1, pp. 59 - 69., (2013).
3. Jae-Hoon Oh, Jae-Hoon Kim, Jong-Woo Kim, .: A Study on the Development of Realtime Online Maketing System Using Web Log Analytics: Society for E-business Studies, Vol.16, No. 3, pp. 249 – 261., (2011).
4. Seok-Hun Kim, Eun-Soo Kim, .: Personalized Advertisement Service Method Using Web Log Mining, The Korean Association of Computer Education, vol.8, No.1, pp. 117-127, (2005).
5. Young-Jik Kwon, Goeng-Wi Jang, .: Web Log Analytics Models : Korea Society of Industrial Information Systems, pp. 212-218, (2009).
6. I. H. Lee, et. al, .: Cst-tree : Cache Sensitive T-trees: Verlag Berlin Heidelberg, (2007)

 Copyright © 2015 SERSC